



Online Onderzoeksomgevingen in 2027

Virtual research
environments (VRE)
over vijf jaar



Op weg naar toegankelijke en laagdrempelige virtuele werkruimten, waarbinnen FAIR, veilig, schaalbaar, automatiseerbaar en over instellingsgrenzen heen kan worden samengewerkt.

Inhoud

1	Managementsamenvatting.....	4
2	Inleiding.....	8
2.1	Aanleiding en doel van dit document	8
2.2	Doel van online onderzoeksomgevingen	8
3	Visie op online onderzoeksomgevingen	10
3.1	Drie scenario's	11
3.2	De visie voor 2027	14
3.3	Workspaces	15
3.4	Federatief werken	18
3.5	Technische architectuur	20
3.6	Organisatorische structuur.....	25
4	Rollen	31
4.1	Algemeen.....	31
4.2	Instellingen en onderzoeksgemeenschappen.....	32
4.3	SURF.....	34
4.4	Marktpartijen en open source gemeenschappen.....	37
	Appendix A.....	39
	Definities	39
a)	Onderdelen	39
b)	Rollen	41
c)	Actoren	42
	Met medewerking van	43
	Figuur 1 – De drie scenario's op een rij	11
	Figuur 2 – Illustratie van onderdelen en actoren in een ecosysteem.....	26
	Figuur 3 – Vereenvoudigd overzicht van onderdelen en actoren	40
	Figuur 4 – Schematisch overzicht van de belangrijkste rollen	41

1 Managementsamenvatting

Met de voortschrijdende digitalisering in wetenschappelijk onderzoek is data-analyse een belangrijk onderdeel geworden van de dagelijkse onderzoekspraktijk. Samenwerking binnen het eigen vakgebied of multidisciplinair, binnen of buiten de grenzen van een instelling is een andere al langer bestaande trend om beter gebruik te kunnen maken van elkaars kennis en sneller tot resultaten te komen.

Groeiende vraag naar veilige data-omgevingen

De combinatie van deze twee ontwikkelingen heeft geleid tot een groeiende vraag naar veilige data-omgevingen voor bijvoorbeeld het delen of hergebruiken van elkaars datasets en algoritmen. Niet langer Excel-sheets naar elkaar mailen, maar een veilige en controleerbare manier om gemakkelijk data te verwerken binnen en tussen instellingen. Meer concreet vragen onderzoekers om virtuele werkruimten, zogenaamde workspaces, die laagdrempelig in het gebruik zijn en waarbinnen ze FAIR, veilig, schaalbaar en automatiseerbaar kunnen samenwerken, over instellingsgrenzen heen. Hiervoor zijn organisatorische en technische faciliteiten en een afsprakenstelsel nodig.

Visie op VRE-ecosysteem

We zien daarom in 2027 een VRE-ecosysteem van portalen en platformen die voldoen aan (inter)nationale standaarden zodat onderzoekers gemakkelijk in hun workspaces kunnen samenwerken met goede toegang tot data en software in repository's. Hiervoor is nodig dat SURF de landelijke coördinatie van het ecosysteem op zich neemt, o.a. door met de leden de bestaande VRE-doelarchitectuur verder door te ontwikkelen, federatief samenwerken te stimuleren en geschikte software licenties af te sluiten.

Het VRE-ecosysteem moet zo ingericht worden dat innovatie van onderdelen goed mogelijk blijft, met een proces om nieuwe diensten in de architectuur te integreren en de mogelijkheid om nieuwe open source ontwikkelde workspaces te delen.

Gekoppelde componenten

Binnen de workspaces moeten benodigde componenten als datasets, software en rekenkracht gekoppeld kunnen worden op basis van technische standaarden (bijvoorbeeld voor koppelvlakken) en op basis van organisatorische afspraken (bijvoorbeeld over rollen van partijen). Het ecosysteem is dus meer dan alleen een technische infrastructuur. De workspaces zullen veelal toegankelijk zijn vanuit

een portaal, bijvoorbeeld van een onderzoekscommunity of de instelling van de principal investigator.

Verskillende gebruikerswensen

De landelijke variëteit in gebruikerswensen zal groot zijn. Er zullen meerdere platformen zijn die workspaces aanbieden, waarbij ieder platform zich richt op specifieke functionaliteiten en expertise. Vanuit de workspaces kunnen vervolgens gemakkelijk precies die data en software worden gebruikt die nodig zijn voor het betreffende onderzoek, op een manier die past bij de gebruikersgroep. Het is cruciaal om workspaces te ontwerpen en aan te bieden met het oog op gebruiksgemak voor minder en onervaren gebruikers.

Federatief samenwerken

Binnen het ecosysteem wordt federatief samenwerken actief gestimuleerd. Onderzoekers en instellingen worden aangemoedigd om componenten, zoals datasets en software, aan te bieden voor landelijk gebruik of tenminste gebruik binnen hun onderzoekscommunity. Beheer en onderhoud van de componenten wordt dan georganiseerd zoals de open source community werkt.

Losjes geïntegreerd

Het landelijke ecosysteem moet losjes geïntegreerd, loosely coupled, zijn. Dat wil zeggen dat de modules niet volledig los van elkaar staan, want dan 'praten' ze niet met elkaar; gemakkelijk samenwerken wordt dan onmogelijk. Maar de modules mogen ook niet volledig geïntegreerd zijn, want dan ontstaat te weinig bewegingsruimte voor onderzoekers en instellingen. Ook wordt continue vernieuwing van het geheel dan moeizaam of zelfs onmogelijk. Bovendien bestaan er al een aantal (initiatieven tot) portalen en platformen. Deze moeten binnen het toekomstige ecosysteem kunnen functioneren.

Onderwerp van nader onderzoek – en discussie, bv. met data stewards en data privacy officers – is hierbij in hoeverre 'loosely coupled' samengaat met security: in hoeverre kan een 'loosely coupled' omgeving ook een veilige omgeving zijn? En is dat altijd noodzakelijk?

Organisatie

Het ecosysteem is georganiseerd in verschillende onderdelen. De belangrijkste zijn portalen en platformen, aansluitend op data-repository's en software-repository's. Portalen bieden toegang en ondersteuning aan gebruikers. Platformen leveren de workspaces, inclusief compute-faciliteiten. Data- en software-repository's zijn opslagplaatsen van respectievelijk data en software en verzorgen directory's om die vindbaar te maken. Binnen het ecosysteem zijn

afspraken gemaakt over rollen voor elk van deze onderdelen. De belangrijkste rollen zijn die van afnemer, leverancier, regisseur en broker. De rollen kunnen op verschillende niveaus in het ecosysteem bestaan, bv. landelijk en binnen een instelling.

Onderzoekers hebben de rol van afnemer (lees: gebruiker), al dan niet georganiseerd in een onderzoeksgemeenschap of instelling. Onderzoeksgemeenschappen en instellingen zullen vaak regisseur zijn van hun portaal en regisseur en broker willen zijn van een eventueel eigen platform. DCC's binnen en buiten de instellingen zullen een belangrijke rol hebben in het VRE-ecosysteem. Data-eigenaarschap en aansprakelijkheid bij datalekken verdienen bijzondere aandacht. Individuele onderzoekers of onderzoeksgroepen, de open source community, instellingen en publieke en private partijen zullen zich opwerpen als aanbieder van componenten en platformen; mogelijk gaat big tech hier een rol spelen, in goede afweging van de publieke-waardendebat. SURF krijgt een coördinerende en aanjagende rol binnen het landelijke ecosysteem, bijvoorbeeld door tot overeenstemming over technische standaarden en de naleving daarvan te komen.

Technologie

Het VRE-ecosysteem bestaat technisch gezien uit componenten en koppelvlakken. De architectuur sluit waar mogelijk aan bij (internationale) referentiearchitecturen. Sommige standaarden voor koppelingen zullen landelijk of zelfs internationaal mogelijk (moeten) zijn, andere standaarden kunnen zich beperken tot bijvoorbeeld een onderzoeksdomein. SURF kan koppeling verzorgen met landelijke voorzieningen, zoals HPC Cloud, Research Drive, toegang tot commercial cloud via SURFcumulus en bijvoorbeeld single sign-on met SURF Research Access Management. Licenties zijn een groot zorgpunt en veelvuldig praktisch obstakel, dat bijzondere aandacht vereist.

Aansluiting bij internationale standaarden en initiatieven is cruciaal. Expliciet worden onder andere veel genoemd: EOSC, European Open Science Cloud en IDS, International Data Spaces.

Gebruikersondersteuning

Een belangrijk aspect binnen het ecosysteem, dat vaak te weinig aandacht krijgt, is gebruikersondersteuning voor de onderzoekers. Het uitgangspunt is dat ondersteuning bij de dienst zit die de betreffende expertise heeft. Dit kan inclusief een onderzoeksondersteunende groep zijn bij een instelling. Schaal, specialisatie en schaarste kunnen een argument zijn om onderdelen van de ondersteuning landelijk te organiseren.

Rollen

In Appendix A zijn *onderdelen*, *rollen* en *actoren* gedefinieerd. Instellingen, onderzoekscommunity's en DCC's zullen een belangrijke rol hebben als regisseur van hun eigen portaal en hun eventuele platform. Gebruikersondersteuning wordt bij voorkeur zo dicht mogelijk bij de gebruikers uitgevoerd. SURF zal een prominente coördinerende rol hebben in de regie en brokerage van het landelijke ecosysteem. SURF zal een voortrekkende rol hebben in de totstandkoming van landelijke standaarden en technische architectuur. Alle betrokken partijen, inclusief marktpartijen waaronder big tech, kunnen een rol hebben als leverancier van datasets, software, componenten, diensten, infrastructuur of zelfs volledige platformen.

Tenslotte: geen eindpunt, maar startpunt!

Wellicht ten overvloede benadrukken we hier dat dit visiedocument geen eindpunt maar *nu een startpunt* zou moeten zijn van:

- Verdere dialoog binnen – en buiten – de SURF-gemeenschap
- Een implementatieprogramma voor een landelijk VRE-ecosysteem

Als uitgebreide samenvatting van dit stuk en weergave van de visie van SURF op online onderzoeksomgevingen is tevens verschenen: De visie van SURF op online onderzoeksomgevingen
<https://www.surf.nl/media/10852>

2 Inleiding

Dit hoofdstuk beschrijft de aanleiding voor en het doel van dit document en het doel van online onderzoeksomgevingen.

2.1 Aanleiding en doel van dit document

In het *SURF 2Jarenplan 2021-2022* is deze ambitie geformuleerd: **Organiseren van een visie op DRE/VRE, met bestaande en/of te ontwikkelen diensten van leden, onderzoeksgemeenschappen, marktpartijen en/of SURF.** Dit document is een resultaat hiervan.

Met dit document willen we de volgende doelen bereiken:

- Een visie schetsen over hoe online onderzoeksomgevingen in Nederland er over vijf jaar (begin 2027) uit zouden kunnen/moeten zien; een visie die de meeste SURF-leden onderschrijven en die een leidraad is voor discussie met en tussen leden van de SURF-gemeenschap de komende jaren.
- In die visie (een begin van) een antwoord geven op de volgende drie vragen:
 - *Ecosysteem* – Moet er een landelijk VRE-ecosysteem komen en in welke mate van ‘geïntegreerdheid’? En zo ja, wat moet er als eerste landelijk gerealiseerd worden:
 - coördinatie op de technische architectuur,
 - coördinatie op federatieve dienstverlening, of
 - coördinatie op inkoop/ontwikkeling van componenten?
 - *Governance* – Welke (bestuurlijke) verantwoordelijkheid moet waar belegd worden? Wie moet welke rol krijgen: wat moeten de onderzoeksgemeenschappen, de DCC’s, de sectoren, de instellingen, SURF en andere partijen doen?
 - *Ondersteuning* – Welke ondersteuning kan waar het beste belegd worden? Wat moeten de onderzoeksgemeenschappen, de DCC’s, researchsupport-afdelingen, ICT-afdelingen, SURF en andere partijen doen?

2.2 Doel van online onderzoeksomgevingen

Voor succes in moderne onderzoek en onderwijs is *samenwerking* een essentiële factor. Zelden publiceert een onderzoeker in zijn eentje, zelden doet een student een college- of afstudeeropdracht in zijn eentje. Vaak werken mensen van dezelfde instelling samen, maar vaker nog werken ze samen met mensen binnen en buiten Nederland, uit andere wetenschappelijke instellingen of uit bedrijven. Dit samenwerken beperkt zich in de praktijk nog vaak tot het online delen of samen schrijven van bijvoorbeeld tekstdocumenten of spreadsheets.

De afgelopen tientallen jaren echter heeft in wetenschappelijk onderzoek de analyse van data een steeds prominentere rol gekregen. Door betere en snellere technologie zijn datasets almaar groter geworden en de verwerking ervan steeds sneller, zeker met de snelle opkomst van machine learning (ML) en andere artificial intelligence (AI) technieken. Maar juist op het vlak van data-analyse werken onderzoekers nog niet veel samen, deels omdat faciliteiten en infrastructures daar niet op ingericht zijn.

Daarom hebben onderzoekers nu een sterk groeiende behoefte aan *digitale/virtuele (onderzoeks)omgevingen*. Samenwerken gaat daarin veel verder dan het uitwisselen van e-mails, tekstdocumenten en spreadsheets. Het gaat om een omgeving waarbinnen onderzoekers op een veilige en correcte manier datasets, algoritmen en onderzoeksresultaten kunnen delen en bewerken (bijvoorbeeld om elkaars algoritmen toe te passen op elkaars datasets of elkaars datasets te combineren).

Online onderzoeksomgevingen zijn expliciet (ook) bedoeld voor samenwerking tussen onderzoekers van *verschillende* instellingen. Dit maakt ze anders dan veel andere IT-toepassingen. Bijvoorbeeld op het gebied van beveiliging van data en standaardisering van koppelvlakken en de toepasselijkheid van softwarelicenties.

3 Visie op online onderzoeksomgevingen

De hier beschreven visie op DRE/VRE voor 2027 is het resultaat van gesprekken en ideeën van vertegenwoordigers van de SURF-gemeenschap over hoe online onderzoeksomgevingen in Nederland *er uit zouden moeten zien*, in een realistische balans tussen *wenselijk* en *haalbaar*. Die balans is belangrijk vanwege de aanzienlijke diversiteit aan wensen en ambities van verschillende (typen) leden van de SURF-gemeenschap.

Over infrastructuur versus ecosysteem

In het SURF 2Jarenplan 2021-2022 wordt gesproken over ‘een federatieve infrastructuur met bestaande en/of te ontwikkelen diensten van leden, onderzoekscommunity’s, marktpartijen en/of SURF. (...) hebben onderzoekers toegang tot lokale, nationale en internationale (Europese) datadiensten en rekenfaciliteiten inclusief public clouddiensten, mits die onder de juiste condities geleverd kunnen worden.’

Hoewel hier gesproken wordt over een infrastructuur, gebruiken we in dit document vooral ook het woord ecosysteem. Veel mensen interpreteren het woord infrastructuur namelijk als de *technische* infrastructuur (de onderste laag/lagen van de architectuur). Een losjes gekoppelde of zelfs integrale ‘infrastructuur’ gaat echter veel verder dan alleen de techniek. Daar spelen ook zaken als architectuur en functionaliteiten, rollen en actoren, financiering en governance een belangrijke rol. En dit is zeker relevant, omdat het 2Jarenplan ook spreekt over een ‘federatieve’ infrastructuur en ‘toegang tot lokale, nationale en internationale (Europese) datadiensten en rekenfaciliteiten’. Dit is eerder ook al in de VRE-referentiearchitectuur beschreven.

Daarnaast straalt het woord ecosysteem meer dynamiek uit dan het woord infrastructuur. En voortdurende innovatie is precies de ambitie voor online onderzoeksomgevingen.

Om deze redenen spreken we liever van een ecosysteem dan van een infrastructuur.¹ Betrokkenen gebruiken overigens ook wel het woord landschap, wat een complexe combinatie van ‘op sommige onderdelen wel en op andere niet geïntegreerd’ suggereert. In dit document zijn de woorden infrastructuur, ecosysteem en landschap grotendeels inwisselbaar. Voor het gemak gebruiken we vooral het woord ecosysteem.

¹ Voor architecten: we hebben het feitelijk over alle architectuurlagen – bv. uit TOGAF 9.0 – en niet alleen over de infrastructuur-gerelateerde laag/lagen.

3.1 Drie scenario's

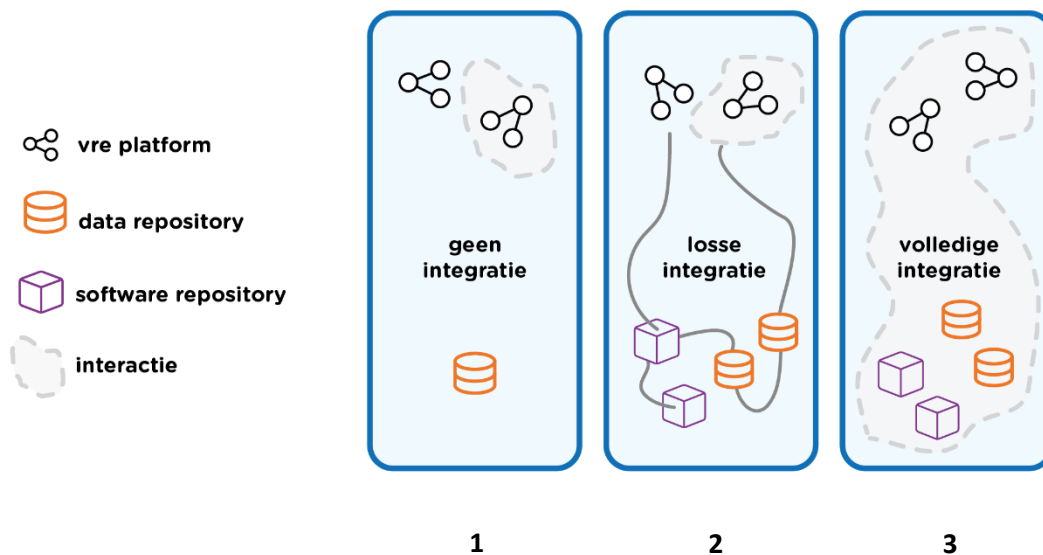
We maken bij het beschrijven van de visie gebruik van drie scenario's:

- ecosysteem van niet-geïntegreerde onderdelen
- losjes geïntegreerd ('loosely coupled') ecosysteem
- volledig geïntegreerd ecosysteem

In discussies met leden van de SURF-gemeenschap kwam het losjes geïntegreerde ecosysteem als meest wenselijke tevoorschijn voor begin 2027.

De drie scenario's lopen op in mate van 'geïntegreerdheid'. Twee belangrijke afwegingen spelen hierbij een rol:

- technische en procedurele *uniformiteit*, waarmee samenwerken en delen voor afnemers gemakkelijk en wellicht goedkoper wordt
- technische en procedurele *diversiteit* waarmee afnemers meer handelingsvrijheid krijgen en waardoor continue vernieuwing gemakkelijker wordt.



Figuur 1 – De drie scenario's op een rij

3.1.1 Scenario 1 – Ecosysteem van niet-geïntegreerde onderdelen

Workspaces worden in dit scenario gerealiseerd door één of meer VRE-platforms. De keuze voor een VRE-platform wordt doorgaans bepaald door de PI, Principal Investigator². Andere deelnemers van het onderzoek sluiten zich hierbij aan. De keuze van het VRE-platform bepaalt de mogelijke diensten, componenten en infrastructuur voor de workspace, en daarmee ook sterk de toepassingsmogelijkheden voor het onderzoeksteam in de workspace. De VRE-platforms zijn dus de facto de bepalende factor in het VRE-ecosysteem.

Sommige instellingen hebben hun eigen VRE-platform. Een aantal commerciële en niet-commerciële partijen bieden mogelijk VRE-platformen aan. De eigenaar/beheerder van een VRE-platform heeft de rol van regisseur en de rol van broker binnen het VRE-platform. Hij bepaalt de regels, en welke diensten, componenten en infrastructuur aangeboden worden op het VRE-platform. Hij koopt in bij en/of geeft toegang aan leveranciers en regelt met hen het beheer en onderhoud.

De technische mogelijkheden van de individuele VRE-platformen bepalen welke, data- en software-repository's koppelbaar zijn vanuit de workspaces, en hoe die koppeling plaatsvindt.

Innovatie van en keuzevrijheid in het VRE-ecosysteem wordt uiteindelijk bepaald door de diversiteit van en de 'concurrentie' tussen VRE-platformen.

VRE-portalen kunnen een keuze maken voor één of meer voorkeursplatformen of aan hun leden hun voorkeur kenbaar maken. Er is geen landelijke regisseurrol of brokerrol.

3.1.2 Scenario 2 – Losjes geïntegreerd ('loosely coupled') ecosysteem

Ook in dit scenario spelen de VRE-platformen een belangrijke rol. Deze rol is echter veel meer in evenwicht met de rol van de VRE-portalen, die ieder in hun eigen domein als vertegenwoordiger van de afnemers fungeren. De VRE-portalen zijn in voortdurende dialoog met de VRE-platformen over principes voor het effectief functioneren van de workspaces, zoals laagdrempeligheid, voldoen aan FAIR-principes, veiligheid, koppelbaarheid, schaalbaarheid en automatiseerbaarheid. Dit leidt tot (domeinspecifieke) afspraken waarvan enkele ook kunnen fungeren als landelijke afspraken. De landelijke regisseurrol en brokerrol wordt ingevuld door een regiegroep en eventueel een landelijke brokergroep. Die bestaan uit

² Principal Investigator: doorgaans de belangrijkste persoon voor de toewijzing van de financiering van een onderzoeksproject

vertegenwoordigers van onderzoeksgemeenschappen en leden van de SURF-gemeenschap. Deze groepen hebben beperkte zeggenschap, zoals het geven van aanbevelingen over federatief werken, referentiearchitectuur en publieke waarden.

Elementen van dit scenario worden hieronder uitvoeriger beschreven.³

3.1.3 Scenario 3 – Volledig geïntegreerd ecosysteem

In dit scenario zijn de onderdelen nog verder geïntegreerd. De landelijke regisseurrol en brokerrol worden ingevuld door een speciaal daarvoor ingerichte organisatie of organisatieonderdeel. Deze organisatie moet ervoor zorgen dat alle onderdelen van het landelijke VRE-ecosysteem voldoen aan de landelijk opgestelde regels en richtlijnen voor gebruiksprincipes en technische architectuur. Het doel daarvan is om alle onderdelen van het VRE-ecosysteem probleemloos voor de afnemer koppelbaar te laten zijn vanuit de workspaces. Ook borgt deze organisatie probleemloze koppeling buiten Nederland op basis van internationale standaarden.

De koppelbaarheid met niet bij het landelijke VRE-ecosysteem aangesloten onderdelen is wellicht mogelijk, maar niet gegarandeerd probleemloos. Vanuit het VRE-ecosysteem kan bijvoorbeeld een richtlijn opgesteld worden (“rules of participation”) waarin wordt beschreven onder welke voorwaarden (technisch, governance, etc) aangesloten kan worden.

3.1.4 Voorkeursscenario

In meerdere discussies met leden van de SURF-gemeenschap is door de leden de voorkeur uitgesproken voor het tweede scenario, een losjes geïntegreerd (‘loosely coupled’) ecosysteem – *overigens, zonder dat dit scenario uitputtend is gedefinieerd*. Belangrijkste twee argumenten zijn dat men (i.) meer vrijheid en mogelijkheid tot vernieuwing wil dan in het volledig geïntegreerde scenario en (ii.) meer compatibiliteit dan in het niet-geïntegreerde scenario. ‘Losjes geïntegreerd’ is dus veel meer het intuïtieve compromis tussen ‘volledig niet geïntegreerd’ (waarbij digitaal samenwerken wel heel erg lastig wordt) en ‘volledig wel geïntegreerd’ (waarbij nauwelijks ruimte is voor eigen interpretatie en innovatie van het geheel). De

³ Overigens is er verschil tussen ‘losjes geïntegreerd’ en ‘loosely coupled’: maximaal geïntegreerde systemen kunnen bestaan uit volstrekt loosely coupled componenten. Een voorbeeld is de auto-industrie: auto’s moeten een volledig geïntegreerd systeem zijn (er is bijvoorbeeld een directe afhankelijkheid tussen het functioneren van het stuur en de wielen). Tegelijkertijd worden auto’s geassembleerd met loosely coupled componenten die in veel gevallen zelfs voor meerdere merken worden toegepast (hetzelfde stuur en dezelfde wielen kunnen bijvoorbeeld op verschillende auto’s worden gemonteerd). Dit soort hergebruik van componenten lijkt zeer wenselijk in het VRE-ecosysteem.

precieze invulling van dit scenario zal dus de komende tijd onderwerp van gesprek zijn binnen de SURF-gemeenschap.

3.2 De visie voor 2027

Niemand kan voorspellen hoe de ontwikkeling van online onderzoeksomgevingen de komende vijf jaar zal verlopen. Daarom lijkt het verstandig dit document jaarlijks aan te passen aan voortschrijdende inzichten.

Het is belangrijk om te benadrukken voor wie we dit doen: voor de onderzoekers die in hun onderzoek gemakkelijk willen kunnen samenwerken met andere onderzoekers. De essentie is dus de workspace, de functionaliteit en het gebruiksgemak daarvan. Technische en organisatorische zaken zijn nodig om de workspaces goed te kunnen laten functioneren.

We verwoorden de visie als volgt:

We zien in 2027 een VRE-ecosysteem van portalen en platformen die voldoen aan (inter)nationale standaarden zodat onderzoekers gemakkelijk in hun workspaces kunnen samenwerken met goede toegang tot data en software in repository's. Hiervoor is nodig dat SURF de landelijke coördinatie van het ecosysteem op zich neemt, o.a. door met de leden de bestaande VRE-doelarchitectuur verder door te ontwikkelen, federatief samenwerken te stimuleren en geschikte software licenties af te sluiten.

Het VRE-ecosysteem moet zo ingericht worden dat innovatie van onderdelen goed mogelijk blijft, met een proces om nieuwe diensten in de architectuur te integreren en de mogelijkheid om nieuwe open source ontwikkelde workspaces te delen.

De manier om adequate workspaces te kunnen bieden en tegelijk continue vernieuwing van het ecosysteem mogelijk te maken is door een losjes gekoppeld ('loosely coupled') landelijk ecosysteem te realiseren: de optimale balans tussen enerzijds *voldoende koppeling(smogelijkheden)* via de workspaces tussen onderdelen zoals portalen, platforms, data-repository's en software-repository's en anderzijds *voldoende flexibiliteit* voor innovatie van het ecosysteem.

Waar die balans *precies* ligt zal de komende jaren onderwerp van voortschrijdend inzicht zijn.

Een speciaal punt van aandacht is *bedrijven* als onderzoekspartner. Steeds vaker zullen bedrijven willen meedoen in de workspaces,

bijvoorbeeld bij onderzoek, maar ook bij bijvoorbeeld afstudeeropdrachten en stages van studenten. Ook dit moet mogelijk worden gemaakt, bijvoorbeeld als het gaat om licenties, security, authenticatie, enzovoort.

Verder is een belangrijk uitgangspunt de lessen die eerder getrokken zijn in het document 'Samenwerkingsmodellen voor VRE-waardeketens'.⁴ De belangrijkste daarin genoemde kritische succesfactoren zijn:

- De aantrekkelijkheid voor afnemers van de geboden workspaces. De workspaces moeten goed tegemoetkomen aan de wensen en eisen van afnemers, inclusief een geschikt serviceniveau, bijvoorbeeld ten aanzien van authenticatie
- De innovatiekracht van het ecosysteem voor online onderzoeksomgevingen, bijvoorbeeld de tijd waarin diensten kunnen worden aangepast aan een gewijzigde realiteit
- Een geschikt licentiemodel voor instellingsoverstijgend gebruik en een passend model voor eigenaarschap van data en de dynamiek van data-delen
- De financiering van de workspaces met alles eromheen moet haalbaar zijn. Het aantal afnemers bepaalt of de aangeboden dienstverlening economisch rendabel kan zijn

3.3 Workspaces

Met workspaces kunnen onderzoekers met iedere gewenste andere onderzoeker gemakkelijk samenwerken op het gebied van data-analyse. Ze kunnen dat doen binnen een aantal randvoorwaarden.

Belangrijke randvoorwaarden geredeneerd vanuit de onderzoeker (de afnemer in het ecosysteem) zijn:

3.3.1 Laagdrempelig

Het gebruiksgemak van een workspace en de gebruikersondersteuning wordt afgestemd op het ervaringsniveau van de afnemer. Afnemers kunnen hun workspace flexibel aanpassen aan hun onderzoekseisen en onderzoekspecifiek maatwerk is mogelijk (inclusief bijvoorbeeld ontsluiting van randapparatuur zoals elektronenmicroscopen). Afnemers ervaren weinig niet-onderzoeksinhoudelijke obstakels (zoals licentiekwesties) en ze ervaren de gebruikskosten als redelijk.

NB. Het is verleidelijk om uit te gaan van de 'persona' van de senior onderzoeker goede vaardigheden op het gebied van data-analyse, automatisering en informatietechnologie. Er

⁴ Zie het document 'Samenwerkingsmodellen voor VRE-waardeketens', versie 1.0, 12 november 2020

zijn ook veel beginnende onderzoekers en studenten met weinig technische kennis en ervaring, aan universiteiten en hogescholen. Een belangrijk uitgangspunt in de visie is daarom, dat er in 2027 verschillende typen workspaces aangeboden worden voor afnemers met uiteenlopende ervaringsniveaus en gebruiksbehoeften. Veel vertegenwoordigers van leden van de SURF-gemeenschap pleiten er zelfs expliciet voor om de focus binnen het landelijke VRE-ecosysteem vooral te leggen bij de minder ervaren gebruikers en zelfs juist ook de 'first time users'. "De experts redden zich wel" is een veelgehoord commentaar. Gebruiksvriendelijkheid is key!

Bij een aantal umc's wordt bijvoorbeeld gewerkt met de volgende vier persona's met verschillende gebruiksbehoeften:

- 1. Onderzoeker die met patiënten werkt**
Werkt met patiëntgegevens, MRI-data, surveys, etc; gebruikt tools zoals EPIC, SPSS
- 2. Onderzoeker met standaard-faciliteiten**
Werkt met patiëntgegevens, lab-data, externe registraties, demografische data, statistische analyses; gebruikt tools zoals SPSS/R, Adobe CC, LimeSurvey, Open Clinica/Castor
- 3. Onderzoeker met behoefte aan flexibiliteit**
Werkt met lab-data, externe registraties, demografische data, statistische analyses; gebruikt tools zoals R, Adobe CC, LimeSurvey, Open Clinica/Castor, experimentele software
- 4. Non-clinical onderzoeker met geavanceerde ict-wensen**
Werkt met lab-data, genetische data, experimentele data, graph-data, visuele data; gebruikt tools zoals Python, R en heel veel domeinspecifieke (eigen) scripts en executables.

In andere wetenschappelijke domeinen zullen vergelijkbare onderverdelingen van afnemers naar ervaringsniveau bestaan. Punt hier is dat het ervaringsniveau van de eerste en laatste persona in het voorbeeld zo ver uiteenlopen, dat ze niet met dezelfde workspace kunnen werken. Ofwel de ene, ofwel de andere zou dan met als vervelend (tot onoverkomelijk) ervaren obstakels geconfronteerd worden. Om twee uiterste voorbeelden te noemen over zowel de technische inrichting van de workspace als de ondersteuning eromheen:

- Voor onervaren afnemers is het vaak al lastig om te begrijpen dat een proces binnen de workspace zomaar kan stoppen of niet eens op gang komt, wanneer niet de juiste parameters zijn meegegeven voor bijvoorbeeld reken capaciteit of opslagruimte, of wanneer een licentie niet geldig is. Een sterk gestandaardiseerde workspace voor beginners zou dat dus moeten afvangen dan wel daarvoor moeten waarschuwen. Een ander voorbeeld is de aanwezigheid van een eenvoudig aanklikbare ‘app store’ met beschikbare (data- en software)componenten, diensten en infrastructuren
- Voor zeer ervaren afnemers is het vaak al een hele klus om een stabiele en goed werkende configuratie te realiseren. Dit vraagt al hun focus. Dan zou het – ‘in the heat of the moment’ – heel fijn zijn, als ze bij een vraag niet eerst door een tijdrovende eerstelijns-hulpdeskprocedure moeten, maar direct de zeer gespecialiseerde tweedelijns-expert kunnen spreken (bellen!). En deze ervaren afnemers zitten helemaal niet te wachten op de beperkingen van een ‘pre-installed app store’, maar willen het liefst zoveel mogelijk vrijheid en regelen het zelf wel.

3.3.2 FAIR, ethisch en veilig

Het beheren en verwerken van de data voldoet aan de FAIR-principes en geldende ethische EU-principes, zoals voor Trustworthy AI.^{5,6} En daarnaast aan de geldende richtlijnen en regels voor veiligheid: in de zin van het voorkómen van diefstal van data en/of intellectueel eigendom (ook van bedrijven als onderzoekspartner), maar ook voor bescherming van persoonsgegevens (GDPR).

3.3.3 Koppelbaar

Vanuit de workspace is het eenvoudig om koppelingen⁷ aan te brengen naar en tussen vooral:

- *datasets* (van onderzoekspartners en uit data-repository’s);
- *algoritmen* (van onderzoekspartners en uit software-repository’s);
- *rekenfaciliteiten* (zoals de nationale supercomputer, maar ook bijvoorbeeld de keuze tussen CPU en GPU);
- *diensten* (zoals cloudopslag of bijvoorbeeld steun bij het opstellen van een data-managementplan);
- *componenten* (zoals scripts en andere software);
- *randapparatuur* (bijvoorbeeld een elektronenmicroscop).

⁵ <https://www.go-fair.org/fair-principles/>

⁶ <https://ec.europa.eu/futurium/en/ai-alliance-consultation.1.html>

⁷ We gebruiken hier de abstracte term ‘koppelingen’ om niet direct in architectuurdiscussies zoals ‘compute-to-data’ te belanden.

3.3.4 Schaalbaar en automatiseerbaar

Indien gewenst kan de workspace opgeschaald worden voor gebruik van grote datasets en grote rekenkracht. De workspace kan modulair worden opgebouwd met componenten op basis van een consistente architectuur. En de workspace ondersteunt de efficiëntie en integriteit van het onderzoek doordat het proces kan worden geautomatiseerd, vastgelegd en herhaald.

3.4 Federatief werken

De ambitie in het *SURF 2Jarenplan 2021-2022* spreekt van een federatieve infrastructuur. Dat past op een heel natuurlijke manier bij een losjes geïntegreerd ecosysteem. Immers, een gefedereerd systeem kent relatief autonome actoren die rollen op zich nemen op basis van een beperkt aantal regels en wederzijdse afspraken. Daarnaast kent het een architectuur op basis van strikt noodzakelijke connectiviteit (in dit geval van het delen van data, software, diensten en infrastructuur in de workspaces).⁸

Online onderzoeksomgevingen lenen zich meer dan veel andere ict-toepassingen goed voor federatief werken. Immers, de aard van online onderzoeksomgevingen is zo, dat afnemers van de ene instelling in de workspace onderdelen van andere organisaties gebruiken en dus ongemerkt al federatief samenwerken. Ook bij DCC's en ict-afdelingen van instellingen groeit de bereidheid om bijvoorbeeld componenten van elkaar te hergebruiken, onder andere omdat ict-budgetten steeds krappere worden voor instellingen. En ook individuele onderzoekers willen vaak componenten hergebruiken die andere onderzoekers hebben ontwikkeld.

Federatie kan plaatsvinden op verschillende onderwerpen, vooral:

- *Datasets* – Het delen van elkaars datasets bijvoorbeeld via verschillende data-repository's
- *Software* – Het delen van elkaars software-modules, applicaties en scripts, bijvoorbeeld via software-repository's

⁸ Relevant in dit kader is dat ook de Nederlandse overheid in haar recent gepubliceerde datastrategie kiest voor een federatief datastelsel (op basis van de reeds bestaande gemeentelijke basisadministraties), gebaseerd op een 'compute-to-data'-architectuur: 'Het federatieve model laat opslag en beheer van data lokaal aan bronhouders, maar faciliteert datagebruik over bronnen heen, centraal door een interoperabel systeem van afspraken en oplossingen op ontsluiting, toegang, annotatie en koppeling. Deze centrale elementen zijn essentieel in het onderscheid tussen een federatief stelsel en een decentraal oerwoud.' Zie voor details:

<https://www.rijksoverheid.nl/documenten/rapporten/2021/10/31/nl-digitaal-interbestuurlijke-datastrategie-nederland>

- *Diensten, componenten en infrastructuur* – Het in workspaces aan elkaar beschikbaar stellen van diensten, componenten en infrastructuur van bijvoorbeeld een instelling, of het bijdragen aan een nationale infrastructuur

Een vraag die nog beantwoord moet worden is wie de verantwoordelijkheid heeft voor beheer en onderhoud van federatief geleverde componenten. Uitgangspunt is dat iedereen die een component aanlevert ook zelf verantwoordelijk is voor beheer en onderhoud, tenzij afgesproken is dat een andere partij die verantwoordelijkheid krijgt. Dit kan bijvoorbeeld de platformleverancier, een open source gemeenschap, een marktpartij of SURF zijn.

3.4.1 Federatief werken in enge zin

In enge zin betekent 'federatief', dat leden van de SURF-gemeenschap zowel de rol van afnemer als de rol van leverancier binnen workspaces op zich kunnen hebben. Het eenvoudigste concept is dat onderzoeker X van instelling A een script heeft geschreven dat vervolgens als component in een workspace kan worden gebruikt door onderzoeker Y van instelling B, of dat de Research Support-afdeling van instelling A een handige component heeft ontwikkeld voor hun eigen onderzoekers die vervolgens ook kan worden afgenomen in workspaces buiten instelling A om.

De rol van afnemer (gebruiker) is evident; daar is het allemaal om te doen. De rol van leverancier betekent, dat leden van de SURF-gemeenschap diensten en componenten leveren die gebruikt worden door:

- afnemers van hun eigen en andere instellingen binnen de SURF-gemeenschap
- afnemers van instellingen en bedrijven (als onderzoekspartner) binnen en buiten Nederland.

Als leverancier van een dienst, component of infrastructuur neemt de betreffende instelling ook het bijbehorende onderhoud en beheer van die dienst of component voor haar rekening, tenzij hiervoor afspraken zijn gemaakt met een andere partij. Dit kan bijvoorbeeld de eigenaar/beheerder van het betreffende platform of van een relevant portaal zijn.

3.4.2 Federatief werken in brede zin

In brede zin kunnen leden van de SURF-gemeenschap bij federatieve online onderzoeksomgevingen iedere gewenste rol op zich nemen, mits dit door het federatieve collectief zo besloten is. Dit geldt in het bijzonder voor de rollen van regisseur en broker.

Verschillende instellingen hebben al in 2021 de wens uitgesproken om regisseur en/of broker te willen zijn binnen het eigen portaal en het eventuele eigen platform. In het beoogde losjes geïntegreerde

scenario kunnen de regisseur- en brokerrol op nationaal niveau in de vorm van een regiegroep en een brokergroep (mede) bemenst worden door vertegenwoordigers uit de SURF-gemeenschap. De zeggenschap beperkt zich tot het geven van aanbevelingen over zaken zoals federatieve richtlijnen, technische referentiearchitectuur en publieke-waardendiscussie.

NB. Binnen de SURF-gemeenschap is intussen aanzienlijke ervaring opgedaan met federatief werken. Een voorbeeld is de federatieve authenticatie die SURF en de instellingen sinds 2008 gerealiseerd hebben. De ervaringen uit dit traject kunnen belangrijke *do's en don'ts* leveren voor federatief werken bij online onderzoeksomgevingen.

3.4.3 Federatief werken op de open source-manier

In 2027 trekken een aantal leden van de SURF-gemeenschap samen op bij het ontwikkelen van specifieke diensten, componenten en infrastructuur voor online onderzoeksomgevingen. Dit zou bijvoorbeeld kunnen gebeuren door teams van onderzoekers eventueel samen met mensen van DCC's, Research Support- of ict-afdelingen van instellingen en SURF. Er wordt dan gewerkt op de open source-manier van werken, dat wil zeggen zonder centrale aansturing op een collaboratieve manier, gebaseerd op peer review en productie door leden van de community, bijvoorbeeld met gebruikmaking van GitLab. Hier zijn uiteraard grenzen aan. Koppelvlakken zullen bijvoorbeeld centraal moeten worden afgesproken en beheerd. Het kan een rol voor SURF zijn om hierin te coördineren: koppelvlakken definiëren en in de HOSA vastleggen.⁹

NB. In de open source wereld is het echter wel gebruikelijk dat diegene die open source van iemand anders gebruikt in zijn project daar zelf de verantwoordelijkheid over neemt om te valideren of dit ook werkt voor zijn/haar use-case. Dat kan op gespannen voet staan met het uitgangspunt dat iedereen die een component aanlevert ook zelf verantwoordelijk is voor beheer en onderhoud.

3.5 Technische architectuur

De structuur van een landelijk ecosysteem moet vooral flexibel zijn. Zo is continue vernieuwing eenvoudig mogelijk zonder dat het de stabiliteit en functionaliteit eronder lijdt. Daarom is een modulaire, gelaagde structuur nodig. Dat mag niet ten koste gaan van de koppelbaarheid, omdat het dan niet werkt: koppelbaarheid is de

⁹ HOSA = Hoger Onderwijs Sector Architectuur

basis van het functioneren van een workspace. Een onderliggende technische architectuur is dus nodig, maar die moet zoveel mogelijk functionele vrijheid laten voor afnemers en leveranciers.

Er bestaan op dit moment al verschillende initiatieven op het gebied van (technische) architectuur voor online onderzoeksomgevingen, met name 'VRE-referentiearchitectuur' en 'Samenwerkingsmodellen voor VRE-waardeketens'.

3.5.1 VRE-referentiearchitectuur

Sinds 2019 bestaat er op initiatief van SURF een VRE-architectuurteam, dat als doel heeft een VRE-referentiearchitectuur te ontwikkelen. Het team bestaat uit architecten van diverse instellingen uit de SURF-gemeenschap. Begin 2020 is een eerste versie van de VRE-referentiearchitectuur gepubliceerd.¹⁰

De VRE-referentiearchitectuur is bedoeld als uitgangspunt en leidraad bij het opstellen van een VRE-architectuur die voldoet aan de specifieke behoeften van onderzoekers en instellingen.

De VRE-referentiearchitectuur is opgesteld vanuit de perspectieven van een aantal belangrijke stakeholders. De onderwerpen die aan de orde komen zijn de stakeholders zelf, hun zorgen en de volgende views: governance, financiën, onderzoeksproces, workspace, datamanagement en technologie. In een volgende versie komen daar de volgende views bij: servicemanagement, ondersteuning en beheer, security en privacy. De referentiearchitectuur geeft onder andere per view een beschrijving van procesmodellen. De VRE-referentiearchitectuur sluit zoveel mogelijk aan bij andere (internationale) referentiearchitecturen, maar mogelijk ook bijvoorbeeld specifiek Europese ontwikkelingen.

3.5.2 Samenwerkingsmodellen voor VRE-waardeketens

SURF heeft eind 2020 een generiek model gepubliceerd dat beschrijft hoe actoren in verschillende rollen als een soort *demand/supply-chain* (VRE-waardeketen) kunnen samenwerken bij het ontwikkelen, leveren en gebruiken van online onderzoeksomgevingen. Dit model is gebaseerd op vier case study's die tezamen representatief zijn voor veel voorkomende situaties. Anders dan de eerdergenoemde VRE-referentiearchitectuur geeft dit document per use case een beschrijving van de specifieke waardeketen, inclusief beschrijving van actoren en rollen en met specifieke focus op de waardepropositie voor afnemers.¹¹

¹⁰ Zie het document 'Doorontwikkeling van een VRE Referentie-architectuur', 13 februari 2020

¹¹ Zie document 'Samenwerkingsmodellen voor VRE-waardeketens', 12 november 2020

3.5.3 Visie voor architectuur

Bovengenoemde twee architectuurdocumenten vormden al een goed uitgangspunt voor verdere ontwikkeling van de meer generieke referentiearchitectuur en specifiekere waardeketens.

Een belangrijke extra input voor architectuur is nu de kenbaar gemaakte voorkeur van vertegenwoordigers van leden van de SURF-gemeenschap voor het scenario van een losjes geïntegreerd ('loosely coupled') ecosysteem. Wat dit technisch precies betekent voor 2027 zal in voortdurende dialoog tussen leden van de SURF-gemeenschap de komende tijd moeten worden overeengekomen. In generieke zin kunnen al wel een zaken benoemd worden:

- Fundamenteel uitgangspunt voor de architectuur is dat onderzoekers goed kunnen samenwerken, en data, software, diensten en infrastructuur met elkaar kunnen delen.
- Wat betekent een 'losjes geïntegreerd' ecosysteem voor de architectuur?
 - Het ecosysteem mag niet uit (volledig) niet-geïntegreerde onderdelen bestaan. Er moet dus een mate van integratie bestaan tussen tenminste een aantal onderdelen van het ecosysteem. De overweging hiervoor is duidelijk: als je afnemers van workspaces (vanuit verschillende instellingen) probleemloos wil laten samenwerken, dan zul je tenminste op een aantal fundamentele technische punten afspraken moeten maken op verschillende onderdelen met elkaar te kunnen laten samenwerken ('koppelvlakken').
 - Het ecosysteem mag niet volledig geïntegreerd zijn. Het kan dus niet zo zijn, dat alle onderdelen naadloos onderling met elkaar geïntegreerd zijn. Los van de vraag of dit ooit organisatorisch haalbaar zou zijn, is de overweging hiervoor duidelijk: op landelijke schaal zou het risico van verstarring van het systeem groot zijn, waardoor innovatie zou worden gehinderd. Daarnaast is er nog een andere belangrijke overweging bij veel instellingen: de behoefte om zelf in voldoende vrijheid vorm te kunnen geven aan de workspaces om de eigen onderzoekers optimaal te kunnen bedienen.
 - In hoeverre kan een 'loosely coupled' ecosysteem ook een veilige omgeving zijn met betrekking tot security (bijvoorbeeld data-diefstal) en privacy? Dit zal zeker niet in alle gevallen samengaan. Wellicht moet er worden gekozen voor workspaces of platforms die heel gebruiksvriendelijk zijn en andere die heel veilig zijn.
- Een voor de hand liggende manier om een goede balans te kunnen vinden tussen niet-geïntegreerd en volledig geïntegreerd is om te werken met modulaire componenten. Dit kent tenminste twee fundamentele aspecten:

- *Componenten* – Binnen de workspaces wordt gewerkt met compatibele componenten, een ‘componentenstrategie’: de workspace wordt dan een ‘blokkendoos’ waar onderzoekers zelfstandig ‘blokken’ in kunnen activeren die goed functioneren. Op deze manier ontwerpen onderzoekers hun eigen online onderzoeksomgeving. Platformen kunnen er dan bijvoorbeeld ook voor kiezen om een catalogus aan te bieden met vooraf geconfigureerde standaard-workspaces, bijvoorbeeld met specifieke functionaliteiten en datasets van onderzoeksgemeenschappen, instellingen en marktpartijen. SURF kan nieuwe en bestaande bouwblokken aanbieden, zoals toegang tot Snellius en LISA, Research Drive, SURF Cumulus, etc.
- *Koppelingen* – In technische zin betreft dit de benodigde interfaces voor het goed in onderling verband kunnen functioneren van de verschillende componenten, binnen en buiten de workspace. Daarnaast betreft dit ook diensten, componenten en infrastructuur die het goed functioneren van de workspaces mogelijk maken, standaarden voor koppelvlakken, zoals API’s; te denken valt ook aan landelijke diensten en infrastructuur voor identity & accessmanagement, zoals SURF Research Access Management (SRAM) en cloudfaciliteiten van marktpartijen. In de minder technische zin gaat dit ook over compliance voor security en privacy en bijvoorbeeld landelijke inkooparrangementen met marktpartijen.
 - NB. Het is essentieel om te begrijpen en ermee akkoord te gaan, dat in een losjes gekoppeld ecosysteem betrokken actoren een zelfstandige verantwoordelijkheid hebben voor de *realisatie* van de koppelvlakken binnen hun eigen onderdeel van het ecosysteem; het is noodzakelijk dat er landelijk afspraken worden overeengekomen (en opgevolgd, uiteraard) over betreffende architectuurprincipes.

De architectuur moet dus zo flexibel zijn, dat vernieuwing goed mogelijk is, ook op de lange termijn. Tegelijkertijd moeten er een aantal dwingende architectuurprincipes gelden, die compatibiliteit en de goede werking van onderlinge onderdelen garanderen.

Een (veel genoemd) *mogelijk* na te streven architectuurprincipe is ‘code-to-data’ of ‘data-bij-de-bron’: data en software blijven zoveel mogelijk op hun oorspronkelijke plek in resp. data-repository’s en software-repository’s en worden vanuit de workspaces met

algoritmen en software benaderd voor verwerking. Er zijn echter nog wel een aantal praktische en meer fundamentele bezwaren die eerst opgelost moeten worden voordat dit vlekkeloos kan gaan werken.

De architectuur van het ecosysteem zou zo gekozen kunnen worden, dat data binnen of buiten een platform of portaal kunnen staan en benaderbaar zijn vanuit de individuele workspaces. De data worden dan dus niet binnengehaald in de workspaces (er worden geen kopieën gemaakt van datasets), maar de applicaties en softwarecode worden naar de data gebracht, dat wil zeggen op de originele data toegepast. Een belangrijke drijfveer is, dat organisaties steeds minder bereid zijn (kopieën van) hun data af te staan; ook voorkomt dit dat met verouderde kopieën van de data wordt gewerkt. Gebruikers willen met live data kunnen werken, niet met kopieën van databases en niet met oude data. Gebruikers van de ene instelling moeten kunnen werken met data van iemand in een andere instelling zonder dat die eigenaar daar last van heeft.

De data in de bron laten staan levert op dit moment nog veel complicaties op. Bijvoorbeeld gebruikt niet iedereen dezelfde metadata, datasets zijn dus vaak niet compatibel. Maar nog fundamenteeler is het volgende punt. Om de echte benefits van 'big data' te bereiken (onverwachte patronen ontdekken) moet je juist heel veel data bij elkaar stoppen, bijvoorbeeld in een *data lake*; dat is heel lastig als je alle data op zijn oorspronkelijke plek laat staan.¹² Ook zou 'data-bij-de-bron' geen dogma mogen zijn; het verantwoord en gecontroleerd uitgeven van data is belangrijker dan het niet delen van data. Immers, als je geen data deelt, kun je ook niet samenwerken.

Een belangrijke vraag in het scenario 'losjes gekoppeld' is in ieder geval de volgende: moet er iemand datastructuren vastleggen, en zo ja waar ligt die taak; is dat een nationale of internationale taak, ligt die taak bij de onderzoeksgemeenschap of ergens anders?

In het algemeen zal de architectuur 'bovenin de stack' meer gefragmenteerd zijn naar gebruikerswensen per doelgroep

¹² Nu zijn ook hier wel weer technische oplossingen voor bedacht, zoals een '(distributed) data mesh', maar die staan vaak nog in de kinderschoenen. Bijvoorbeeld, met data mesh wordt binnen individuele organisaties wel geëxperimenteerd, maar er lijken op dit moment geen geslaagde implementaties van te zijn over de grenzen van organisaties heen. Dit zou per wetenschappelijk domein een rol kunnen zijn voor onderzoeksgemeenschap. En er zijn ook andere, meer generieke initiatieven voor data-delen tussen organisaties, zoals bij de NL-AI Coalitie en Gaia-X IDS (International Data Spaces), maar ook die staan nog in de kinderschoenen.

(onderzoeksteams, onderzoekscommunity's, instellingen) en 'onderin de stack' zoveel mogelijk uniform en nationaal (netwerken, hardware, autorisatie/authenticatie, etc). Er zullen in enige mate 'onderliggende' compatibiliteitsstandaarden moeten worden afgesproken om de verschillende onderdelen van het ecosysteem met elkaar te kunnen laten functioneren. Tegelijk moet er ook worden afgesproken wat vrijgelaten wordt voor de gebruikers.

Anders gezegd: de technische architectuur is deels 'inzichtgevend' en deels 'richtinggevend'. Naarmate we meer onderin de technische stack komen zal de technische architectuur meer richtinggevend zijn, en meer bovenin de stack zal die meer inzichtgevend zijn. De komende tijd zal gezamenlijk moeten worden vastgelegd aan welke technische standaarden iedereen in het ecosysteem zich zal moeten conformeren om effectief mee te kunnen doen, welke voorkeursstandaarden er zijn waar eventueel van kan worden afgeweken en over welke technische terreinen geen bindende afspraken bestaan.

Het VRE-referentiearchitectuurteam moet een belangrijke rol krijgen bij het vastleggen van de uiteindelijke architectuur voor het landelijke VRE-ecosysteem.

3.6 Organisatorische structuur

De voorkeur voor het scenario van losjes geïntegreerd ('loosely coupled') ecosysteem heeft belangrijke consequenties voor de organisatie.

Uitgangspunt is – zoals steeds – dat de afnemers van de workspaces probleemloos kunnen werken zonder dat de mogelijkheden tot continue vernieuwing van de workspaces in gevaar komen.

Dat gebeurt door de technologie 'losjes gekoppeld ('loosely coupled')' in te richten en om dat te kunnen bereiken is weer een bepaalde vorm van organisatie nodig. Dat is de fundamentele volgorde: *afnemers* → *workspaces* → *technologie* → *organisatie* (de afnemers hebben wensen over de inrichting van de workspaces; de inrichting van de workspaces bepaalt de benodigde technologie en integratie; de technologie bepaalt welke mate van organisatie en afstemming nodig is).

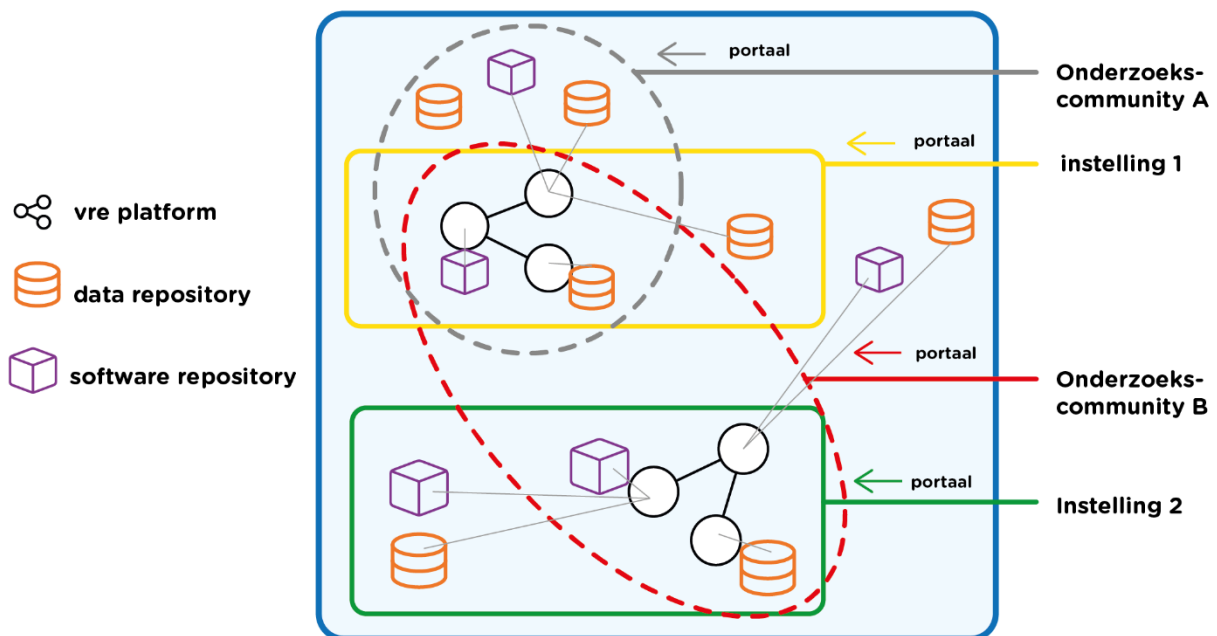
3.6.1 Visie voor organisatie

'Losjes gekoppeld' betekent dat er in 2027 een beperkte mate van koppeling bestaat. Omdat onderdelen van het ecosysteem door verschillende actoren, mogelijk in verschillende rollen, worden beheerd, betekent dit dat – waar dit relevant is – actoren hun activiteiten en infrastructuur aan elkaar willen koppelen. Dit

betekent dat er technische en organisatorische (en eventueel financiële) afstemming moet plaatsvinden plus de uitvoering van gemaakte afspraken. Hiervoor is een vorm van organisatie nodig.

Een belangrijke organisatorische taak is bijvoorbeeld ook een (landelijke en/of domeinspecifieke) catalogus/catalogi waarmee onderzoekers kunnen achterhalen waar ze welke datasets kunnen vinden.

Om één en ander te kunnen toelichten is het belangrijk eerst een beeld te hebben van de betreffende onderdelen. We doen dit aan de hand van het voorbeeld in de figuur hieronder:



Figuur 2 – Illustratie van onderdelen en actoren in een ecosysteem

We gaan uit van een fictieve situatie met twee VRE-platformen X en Y (dat zouden op dit moment bijvoorbeeld anDREa en SURF Research Cloud kunnen zijn), twee instellingen 1 en 2 (bijvoorbeeld een universiteit en een hogeschool), twee onderzoeksgemeenschappen A en B (bijvoorbeeld een health-community en een community van AI-onderzoekers) en diverse data- en software-repository's (zoals nu bijvoorbeeld DANS, EOSC of OpenML).

Hoe werkt het dan in 2027? Het eenvoudige voorbeeld is dat van onderzoeksgemeenschap A. In dit geval hebben alle onderzoekers een aanstelling bij instelling 1 (wat dus eigenlijk *juist niet* de essentie is van online onderzoeksomgevingen). Het portaal van onderzoeksgemeenschap A heeft kennelijk besloten (of beveelt aan) om in zee te gaan met platform X. Leden van onderzoeksgemeenschap A zijn dus afnemers van de workspaces van platform X. Het ligt voor de hand, dat domeinspecifieke ondersteuning geleverd wordt door leden van community A en dat meer technische en organisatorische ondersteuning geleverd wordt door het DCC en/of de ICT-afdeling van instelling 1. In dit eenvoudige voorbeeld is er alleen ontsluiting geregeld naar data-repository's en software-repository's binnen de eigen community A dan wel binnen de eigen instelling 1.

Het complexere, maar relevantere voorbeeld is dat van onderzoeksgemeenschap B. In dit voorbeeld hebben onderzoekers uit deze community een aanstelling bij instelling 1 of bij instelling 2 (of bij beide). Onderzoeksgemeenschap B heeft besloten toegang te faciliteren tot twee verschillende platformen, X en Y. Kennelijk hebben leden van de community uiteenlopende specifieke eisen voor een workspace waar niet door één platform aan voldaan kan worden. Het kan dus gebeuren dat een onderzoeksteam van een onderzoeker uit instelling 1 samenwerkt met een onderzoeker uit instelling 2 op platform Y. Laten we zeggen dat de onderzoeker uit instelling 2 de PI is. Wanneer nu onderzoeker 1 een hulpvraag heeft, dan zijn er twee mogelijkheden:

- Onderzoeksgemeenschap B heeft zelf faciliteiten voor het opstarten en technisch ondersteunen van het gebruik van workspaces ingericht. Dan kan onderzoeker 1 daar terecht.
- Onderzoeksgemeenschap B heeft geen faciliteiten voor het opstarten en technisch ondersteunen van het gebruik van workspaces ingericht. Dan zal onderzoeker 1 ondersteuning vragen bij de instelling van de PI, in dit geval instelling 2.

Verder hebben kennelijk community B en/of platform Y toegang mogelijk gemaakt tot data-repository's en software-repository's buiten instelling 1 en 2 en ook buiten het directe bereik van community B.

Dit zijn maar twee eenvoudige voorbeelden. Toch blijkt hier hopelijk uit hoe snel de complexiteit van een ecosysteem groot kan worden.

Daarom is het ook bij een 'losjes gekoppeld ('loosely coupled')' ecosysteem nodig om van tevoren goed over de organisatie ervan na te denken.

Dus wat betekent 'losjes gekoppeld' voor de organisatie? In ieder geval het volgende:

3.6.2 Minimale vereisten voor koppelbaarheid

Op grond van het uitgangspunt 'afnemers → workspaces → technologie → organisatie' bepalen technische vereisten voor een groot deel de organisatievorm van het landelijk ecosysteem. De componentenstrategie met koppelvlakken vereist dat er een minimale set aan afspraken is over technische standaarden en dat die standaarden ook worden nageleefd: het richting gevende deel van de architectuur.

In de organisatie van het ecosysteem is hiervoor nodig dat er een organisatie of organisatieonderdeel is met voldoende expertise, autoriteit en draagvlak bij alle betrokkenen om totstandkoming en naleving van de minimale set aan afspraken over technische standaarden te garanderen. Hiervoor is een vorm van landelijke *coördinatie* nodig, inclusief monitoring van en waar nodig aansluiting op internationale standaarden. Standaarden evolueren in de loop van de tijd. Dit betekent dat het ecosysteem voortdurend 'in beweging' zal zijn en dat een partij hier regie op moet voeren.

De coördinatie gaat verder dan alleen maar technische standaarden. Het gaat bijvoorbeeld ook over organisatorische en juridische zaken. Te denken van aan coördinatie van bijvoorbeeld overlappende rollen onder actoren binnen het ecosysteem of het evenwichtig balanceren van financiële en niet-financiële kosten en baten. Ook valt te denken aan het organiseren van landelijke inkoop bij marktpartijen en bijvoorbeeld het overeenkomen van licentieovereenkomsten die specifiek geschikt zijn voor online onderzoeksomgevingen.

3.6.3 Minimale vereisten voor flexibiliteit en vrijheid

Naast het confirmeren aan een minimale set van standaarden is er de noodzaak tot flexibiliteit van inrichting van workspaces en de behoefte bij leden van en betrokkenen bij de SURF-gemeenschap aan vrijheid.

De benodigde flexibiliteit bij de inrichting van individuele workspaces leidt tot de behoefte aan een architectuur die innovatie van het ecosysteem bevordert. Dit betekent, dat er regelmatig onderdelen in het ecosysteem zullen moeten kunnen ontstaan en verdwijnen. Een voorbeeld is 'streaming data'; hier is nu nog niet veel vraag naar, maar met de sterke opkomst van bijvoorbeeld digital twins lijkt inrichting van workspaces die 24x7 streaming data kunnen verwerken in de nabije toekomst al onvermijdelijk. Een open

vraag is hoeveel (landelijke) coördinatie op dit soort innovatie gewenst is en in welke vorm.

Leden van de SURF-gemeenschap hebben duidelijk aangegeven behoefte te hebben aan een aanzienlijke mate van eigen vrijheid in de inrichting en ontwikkeling van het ecosysteem. In de organisatie van het ecosysteem betekent dit dat leden van en betrokkenen bij de SURF-gemeenschap voldoende ruimte moeten krijgen om een eigen invulling te kunnen geven aan workspaces en andere onderdelen van het ecosysteem.

Een discussie uit de voorbereiding op dit document die nog niet besloten is, is de vraag wat hierin landelijk de dominante dimensie zou moeten zijn. Anders gezegd, naar welke dimensie zouden we portalen en platformen landelijk vooral willen fragmenteren – als zo'n keuze überhaupt gemaakt gaat worden:

- *Veiligheid van de data* – Sommige workspaces zullen heel veilig moeten zijn en aan de allerhoogste privacy-eisen voldoen, zoals GDPR, FAIR-principes, IP-spionage, datadiefstal, etc.; andere workspaces kunnen worden ingericht voor toepassingen die niet vallen onder GDPR, bijvoorbeeld als er geen persoonsgebonden data zijn, etc.
- *Vaardigheid van de gebruiker* – Er is vraag naar workspaces die ingericht zijn naar het expertise-niveau van de gebruiker; bijvoorbeeld een workspace voor beginners, voor gevorderden en voor experts, inclusief bijbehorend niveau van gebruikersondersteuning
 - Dit speelt met name bij de hogescholen en in onderwijstoepassingen; het liefst willen zij workspaces die 'meegroeien' met de expertise van de gebruiker.
 - In het verlengde hiervan: er is een groot verschil in technische en support-wensen tussen de groep Windows-gebruikers en de groep Linux-gebruikers.
- *Use cases/user journeys* – Sommigen pleiten ervoor om als onderscheidende dimensie te nemen een mengvorm van user journey (aan te geven vanuit de community's) en use case (een technische insteek, bijvoorbeeld een opzet voor Machine Learning of een opzet voor simulatie).
- *Wetenschapsdomein* – Portalen en een aantal platforms zullen qua functionaliteit ingericht zijn naar de wensen van een specifieke onderzoeksgemeenschap.
- *Instelling* – Een aantal instellingen zal hun eigen VRE-portaal willen kunnen aanbieden aan hun eigen medewerkers (als PI).
- *Risico-classificatie* – Een meer juridisch ingestoken fragmentatie van architectuur, datasets en functionaliteit is in termen van risico: wat kan er misgaan en wie heeft op welk moment welke verantwoordelijkheid, wie is probleemeigenaar als het misgaat, hoe kun je risico's compartimenteren?

Vermoedelijk kan niet gekozen worden voor één dimensie, maar zullen deze allemaal in meer of mindere mate een rol spelen. Het ecosysteem zal dus bestaan uit meerdere, wellicht zelfs vele, portalen, platformen, data-repository's en software-repository's. Op dit moment lijkt de dimensie van wetenschapsdomein dominant, met de huidige dominantie van het medische domein.

4 Rollen

Voor online onderzoeksomgevingen zijn in ieder geval de volgende rollen gedefinieerd: afnemer, leverancier, ondersteuner, regisseur en broker (zie Appendix A). Verschillende actoren kunnen verschillende rollen hebben op verschillende onderdelen van het ecosysteem. In dit hoofdstuk inventariseren we voor de hand liggende rollen voor de betreffende actoren.

4.1 Algemeen

In de dialogen met leden van de SURF-gemeenschap is de wens te kennen gegeven in principe elke rol te kunnen laten bestaan in elk onderdeel van het ecosysteem. Eén of meer actoren kunnen dan houder van die rol zijn.

Bijvoorbeeld: Een instelling met een eigen portaal en platform zal regisseur willen zijn van het eigen instellingsportaal en zal regisseur en broker willen zijn van het eigen instellingsplatform. Tegelijk kan er een regisseur en broker worden aangesteld voor het landelijke ecosysteem.

Het is van belang deze rollen dus goed af te bakenen en op elkaar af te stemmen. Verder moet erop gelet worden dat een partij die een rol voor zichzelf ziet, die rol ook voldoende kan uitvoeren: is men er adequaat voor georganiseerd, heeft men de benodigde fondsen en resources?

Uit de huidige stand van zaken van discussie lijkt de volgende aanzet tot rolverdeling een goed uitgangspunt voor verdere discussie tot 2027.

Voor een aantal principes lijkt consensus te bestaan, met name:

- Er kan in 2027 een landelijke ‘regiegroep’ zijn die de rol van regisseur van het landelijke ecosysteem op zich neemt met vertegenwoordigers van leden van de SURF-gemeenschap. SURF kan daarin een coördinerende rol hebben. Zo zou er ook een landelijke ‘brokergroep’ kunnen zijn.
- Er zullen een aantal door (bijna) iedereen gebruikte ‘standaard’ diensten, componenten en infrastructuur zijn. Deze kunnen bijvoorbeeld in een soort standaard ‘referentie-workspaces’ centraal worden aangeboden en beheerd, bijvoorbeeld door SURF, de eigen instelling of een commerciële leverancier. Daarnaast kunnen in de onderzoeksgemeenschap en instellingen bijvoorbeeld een soort DevOps-teams bestaan die specifieke functionaliteiten ontwikkelen en ook het beheer daarvan doen.

- Er zou een landelijk bestand kunnen komen van beschikbare datasets en een landelijke inkoopregeling.

Een aantal onderwerpen vraagt om verdere discussie. Dit betreft bijvoorbeeld:

- Welke diensten, componenten en infrastructuur vallen onder de noemer ‘standaard’ en ‘niet-standaard’ moet nog worden uitgewerkt.
- Er zijn actoren die in dit document niet voldoende belicht zijn, zoals eScience Center en TNO. eScience Center zou bijvoorbeeld veel kunnen betekenen in de rol van ondersteuner en in het bouwen van software-modules. TNO zou een rol kunnen hebben bij het ontwikkelen van open source software.
- Gebruikersondersteuning is een onderbelichte rol. Een aantal instellingen en onderzoeks-community’s willen deze rol graag oppakken, bijvoorbeeld in het DCC, maar voorzien dat er wellicht te weinig resources beschikbaar zullen zijn.
- Veel onderzoekscommunity’s zijn erg decentraal georganiseerd en veel DCC’s zijn nog in oprichting. Het is de vraag of zij in staat zijn zich voldoende goed te kunnen organiseren om hun collectieve vraag goed te articuleren. De thematische DCC’s zouden hier een centrale rol in kunnen gaan pakken.
- De DCC’s kunnen de discussie intensiveren over hun rol binnen een landelijk ecosysteem en over de architectuur en organisatie van hun online onderzoeksomgeving (zowel portaal als evt. platform).
- De rol van Research Support, data stewards etc. bij de instellingen is voor online onderzoeksomgevingen heel belangrijk en wellicht onderbelicht. Het gaat dan vooral over eerstelijns-gebruikersondersteuning en (eventueel) het ontwikkelen van maatwerkoplossingen.
- Financiering en bekostigingsmechanismen van portalen, platforms en het gehele ecosysteem is tot op heden niet besproken en moet verder uitgewerkt worden.

4.2 Instellingen en onderzoekscommunity’s

Instellingen en onderzoekscommunity’s willen zelf de regie voeren over hun portaal of platform. Ook willen ze de eerstelijns-gebruikersondersteuning zelf doen. Hoeveel taken ze zelf willen uitvoeren, verschilt: sommigen willen veel in eigen hand houden, anderen willen veel uitbesteden en zich beperken tot proces-regisseur van ingekochte functionaliteiten. Men is het erover eens dat met name taken rond de landelijke infrastructuur, centraal door een landelijke organisatie moeten worden uitgevoerd.

Over de volgende punten was tijdens de voorbereiding van dit document in redelijke mate consensus onder leden van de SURF-gemeenschap.

- Onderzoeksgemeenschap's (en instellingen) zijn eigenaar van en vormen de portalen.
- Onderzoeksgemeenschap's en instellingen willen zelf tenminste de rol van regisseur en broker uitvoeren voor het portaal van de eigen instelling of onderzoeksgemeenschap. Binnen hun eigen instelling of community kunnen ze zelf:
 - De regie houden op het gebruik en de opzet van het eigen portaal
 - Hun gebruikers begeleiden in het toegang geven tot en gebruiken van de workspaces
 - Registratie van gebruikers en verrekening van kosten bijhouden met een eigen dashboard
 - Ondersteuning aan gebruikers geven bij de inrichting van de workspaces
 - De rol oppakken van 'data broker': het opstellen van een data-dictionary voor meta-data en een catalogus bijhouden van welke data waar te vinden zijn
 - FAIR datamanagement en het opstellen van FAIR data catalogi
 - Bepalen welke leveranciers diensten kunnen aanbieden in de workspaces en de levering van diensten door leveranciers bewaken.
- Instellingen zien een grote rol voor hun eigen DCC bij hun eigen portaal. Het DCC kan fungeren als eerstelijns-ondersteuning voor de gebruikers van de eigen instelling, bijvoorbeeld als data steward of lokaal ontwikkelaar van maatwerkoplossingen.

In principe komen alle mogelijke gebruikerstypen in alle organisaties in de SURF-gemeenschap voor. Toch zijn er belangrijke verschillen, bijvoorbeeld:

- Onderzoeksgemeenschap's en onderzoeksinstellingen zijn puur gericht op (wetenschappelijk) onderzoek en hebben vaak heel geavanceerde gebruikers, bijvoorbeeld vaak met gebruik van Linux en andere open source tools.
- Universiteiten bestaan deels uit onderzoeksgemeenschap's, maar deels ook uit docenten en studenten. Hier is een mengeling van geavanceerde en minder geavanceerde gebruikers, die vaak Windows gebruiken.
- Hogescholen hebben naast onderzoekers ook veel docenten en studenten als afnemers van online onderzoeksomgevingen. Deze groepen hebben vooral behoefte aan een eenvoudige workspace die 'gewoon werkt'. Gebruikersondersteuning is hier heel belangrijk.

- In onderwijstoepassingen wordt bij stages en afstudeeropdrachten vaak samengewerkt met bedrijven. Voor de samenwerking met bedrijven moet dus ook een oplossing aangeboden worden.

Het is dus zeer van belang in het landelijke ecosysteem heel goed na te gaan of alle verschillende doelgroepen worden bediend met workspaces die passen bij hun gebruiksprofiel, van zeer geavanceerde gebruikers tot en met startende gebruikers.

De volgende punten behoeven de komende jaren verder uitwerking:

- De mate waarin instellingen en onderzoekscommunity's de rollen willen invullen verschilt aanzienlijk; sommigen willen juist alles naar zich toe trekken en een eigen platform ontwikkelen, andere willen alleen regisseur zijn van door hen of collectief door SURF ingekochte diensten, componenten en infrastructuur. Het ecosysteem in 2027 zal deze diversiteit moeten kunnen faciliteren.
- Een workspace wordt meestal aangemaakt/aangevraagd door de PI van een onderzoeksteam. Die doet dat logischerwijze bij de eigen instelling. Vervolgens gaan mensen van andere instellingen ook in die workspace werken. Als die vragen hebben, kunnen die dan zomaar terecht bij de support-organisatie van de oorspronkelijke instelling? Vaak niet, is nu nog de praktijk.
- Als er gebouwd moet worden, zijn er dan goede ontwikkelaars beschikbaar? Er is nu al een groot tekort aan bijvoorbeeld technisch architecten en beheerders.

4.3 SURF

Leden van de SURF-gemeenschap zijn het erover eens dat er een rol voor SURF is in het nationale ecosysteem. Die splitst zich uiteen in twee hoofdtaken:

1. Coördinatie van het ecosysteem met betrekking tot de landelijke regie- en brokerrol (dus een governance-rol)
2. Exploitatie van (onderdelen van) het landelijk deel van de infrastructuur voor het ecosysteem.

NB. De mate waarin leden het wenselijk vinden om SURF die taken te laten uitvoeren heeft vooral te maken met de vraag of men SURF *operationeel* in staat acht om de door SURF op zich genomen taken en verantwoordelijkheden ook daadwerkelijk succesvol uit te voeren en niet zozeer met de vraag of SURF *ten principale* de partij zou zijn die hiervoor in aanmerking komt.

Het is aan SURF en de leden om hier in de aanloop naar 2027 gezamenlijk aan te werken.

De meeste onderzoekscommunity's en leden zien een rol voor SURF (als onderdeel van de 'regiegroep', dus samen met leden van de SURF-gemeenschap) als *coördinator* van (vooral de landelijke aspecten van) een landelijk ecosysteem, bijvoorbeeld:

- Het coördineren van de governance van het landelijke ecosysteem
- Afstemmen van architectuurprincipes en gedragsregels als onderdeel van de architectuur en governance van het ecosysteem
- Coördineren dat er voldoende financiering is en blijft voor de landelijke onderdelen van het ecosysteem
- Een rol spelen om het vertrouwen bij onderzoekscommunity's en instellingen in het nationale ecosysteem op te bouwen
- Coördineren dat de instellingen min of meer dezelfde lijn willen volgen en in hetzelfde tempo mee kunnen gaan met de ontwikkelingen van de online onderzoeksomgevingen
- Coördineren van licenties.
- Vernieuwing (innovatie) van het landelijke ecosysteem coördineren en (grotendeels via partners) vormgeven.
- SURF is voor de leden de aangewezen partij voor het coördineren en maken van (landelijke) afspraken over architectuurstandaarden binnen het ecosysteem. Onderwerpen zijn bijvoorbeeld:
 - Requirements opstellen; technische aspecten van de architectuur
 - Licentie-problematiek voor online onderzoeksomgevingen oplossen
 - Afspraken maken rond security en privacy
 - Uniforme afspraken voor datastromen
 - Uniforme verwerkingsovereenkomsten
 - Generieke afspraken over koppelvlakken
 - Compatibiliteit tussen platformen, data-repository's en software-repository's
- SURF kan landelijke inkoop en contractbeheer doen en ervoor zorgen dat licenties geldig zijn binnen het ecosysteem, o.a. onderhandelen met (de grote) marktpartijen namens alle instellingen. Maar niet alleen bij marktpartijen; SURF kan ook een rol spelen bij het 'inkopen' van producten en diensten bij open source community's en individuele instellingen (bijvoorbeeld TNO).
- SURF kan delen van de landelijke infrastructuur voor het ecosysteem (mede) realiseren en beheren. Dit kan bijvoorbeeld gaan over:
 - Landelijke infrastructuur voor autorisatie/authenticatie

- Landelijke infrastructuur voor uitwisseling van data en software
- Beheer van wat binnen het ecosysteem dat onder de noemer ‘nationale infrastructuur’ valt.
- Landelijke uitrol en beheer van door instellingen of onderzoeksgemeenschappen ontwikkelde componenten.
- SURF kan een dienstencatalogus van het landelijke ecosysteem coördineren, uitgeven en beheren, inclusief een kwalificatie van bijvoorbeeld het gebruiksgemak, benodigde licenties, etc. Zeker bij federatief werken is het nodig dat een partij zorgt voor vindbaarheid van door instellingen ontwikkelde componenten; al zou deze rol tenminste deels ook bij de onderzoeksgemeenschappen en instellingen gelegd kunnen worden (maar dan nog is er behoefte aan een landelijke rol, tenminste coördinerend).
- SURF zou beperkt nog producten zelf moeten ontwikkelen, en in plaats daarvan zich meer richten op het stimuleren van instellingen, allianties en onderzoeksgemeenschappen bij het inkopen, beheren en evt. aanpassen en landelijk als dienst ter beschikking stellen van bestaande open source en commerciële producten en diensten, bijvoorbeeld:
 - SURF kan coalities opzetten en bevorderen waarbinnen schaalbaar en repliceerbaar kan worden samengewerkt op basis van eerder gestelde architectuureisen.
 - SURF kan gunstige gebruiksvoorwaarden garanderen, bijvoorbeeld voor eenvoudig gebruik van Microsoft Azure, Google Cloud of Amazon AWS met de juiste authenticatie.

Over de volgende punten was niet echt consensus tussen de deelnemers van de interviews of deze punten behoeven nog uitwerking.

- Moet SURF zelfstandig een eigen landelijke referentiearchitectuur maken voor VRE of grotendeels voortbouwen op internationale architecturen, zoals Gaia-X en IDS (International Data Spaces) en deze ontwikkelingen nauwgezet bijhouden? Misschien moet SURF er dan voor zorgen dat Nederlandse partijen die binnen het ecosysteem actief zijn hierbij aansluiten.
- SURF zou zich met betrekking tot online onderzoeksomgevingen meer moeten opstellen als broker van kennis in plaats van als ontwikkelaar van producten, bijvoorbeeld inzetten op (coördinatie van) educatie van de sterk groeiende niet zo technisch-vaardige data scientists die op weg moeten worden geholpen. De instellingen, zeker de hbo-instellingen, kunnen dat niet alleen.

- Sommige onderzoekers willen direct ondersteuning van experts bij SURF. Maar de DCC's en de ict-afdelingen van meeste instellingen willen vaak dat de supportvraag via de instelling zelf verloopt. Hierover moeten afspraken gemaakt worden. Vraag hierbij is: welke specialismen zou SURF dan moeten aanbieden?

4.4 Marktpartijen en open source community's

4.4.1 Verschillende workspaces

Er is een spanningsveld tussen publieke waarden en big tech. Dit komt vooral tot uiting in het functioneel sterke, gunstig geprijsde aanbod van big tech versus het vaak gratis aanbod vanuit de open source gemeenschap (waar veel onderzoekers zelf actief lid van zijn). Deze 'stammenstrijd' tussen meer principiële en meer pragmatische onderzoekers heeft consequenties voor het ecosysteem. Binnen de onderzoeksgemeenschap in Nederland zijn de twee uitersten Linux-gebruikers en Windows-gebruikers. Voor veel Linux-gebruikers is het gebruik van tools van commerciële bedrijven ongewenst, terwijl veel wat minder ervaren gebruikers alleen de Windows-wereld kennen. Aan deze twee doelgroepen zullen wellicht verschillende soorten workspaces moeten worden aangeboden.

Er zullen bijvoorbeeld vermoedelijk onderzoekers zijn die graag een workspace bouwen met open source tools, die gemakkelijk te gebruiken is. Daarbij zouden commerciële tools zoals SPSS of Matlab vervangen worden door open source equivalenten. Het lijkt in het kader van publieke waarden een goed gesprek waard om te onderzoeken of hierin stappen kunnen worden gezet. Het zou toch mooi zijn als studenten niet automatisch bv. AdobeCC of SPSS gebruiken, maar vanaf dag één leren werken met alternatieven.

4.4.2 (On)afhankelijkheid van big tech

Daarnaast speelt de discussie over publieke waarden: hoe (on)afhankelijk willen we zijn van (vooral Amerikaanse) grote commerciële leveranciers? Ook hierover zijn niet alle leden van de SURF-gemeenschap het eens. Sommigen vinden dat we zoveel mogelijk gebruik moeten maken van open source, aangevuld met eigen nationale ontwikkelingen (zoals SURF Research Cloud). Een andere groep vindt dat we niet te veel functionaliteiten moeten bouwen, omdat de markt (inclusief big tech) die straks waarschijnlijk goedkoper en betrouwbaarder kan leveren. Deze groep zet in op het maken van goede, bindende afspraken met commerciële leveranciers over het voldoen aan de eisen van publieke waarden.

Dat spitst zich vooral toe op eisen rondom veiligheid en privacy, en op het maken van afspraken om *vendor lock-in* te minimaliseren.

Een belangrijke vraag is dus of en op welke manier big tech zich in de komende jaren gaat richten op het aanbieden van (infrastructuur voor) online onderzoeksomgevingen. Deels doet big tech dat overigens al; zo is anDREa ontwikkeld op Microsoft Azure. Dit moet in ieder geval de komende jaren nauwkeurig in de gaten gehouden worden.

In de discussies over een losjes gekoppeld ecosysteem zal dit thema nog veel en intensief besproken moeten worden.

Appendix A.

Definities

We spreken in dit document over ‘online onderzoeksomgevingen’ waar anderen vaak spreken over Digital Research Environments (DRE) of Virtual Research Environments (VRE). Deze termen worden vaak gebruikt voor zowel een ecosysteem of infrastructuur als voor een portaal of platform, als voor een individuele werkruimte. Dan is niet duidelijk wat precies bedoeld wordt met ‘de VRE’. Om spraakverwarring te voorkomen gebruiken we in dit document de definities van onderdelen zoals hieronder gegeven.

Daarnaast worden *rollen* en *actoren* vaak door elkaar gehaald. Daarom definiëren we die ook.

Voor de volgende definities sluiten we aan bij definities in de HORA, Hoger Onderwijs Referentie Architectuur en in het document ‘Samenwerkingsmodellen voor VRE-waardeketens’.^{13,14}

a) Onderdelen

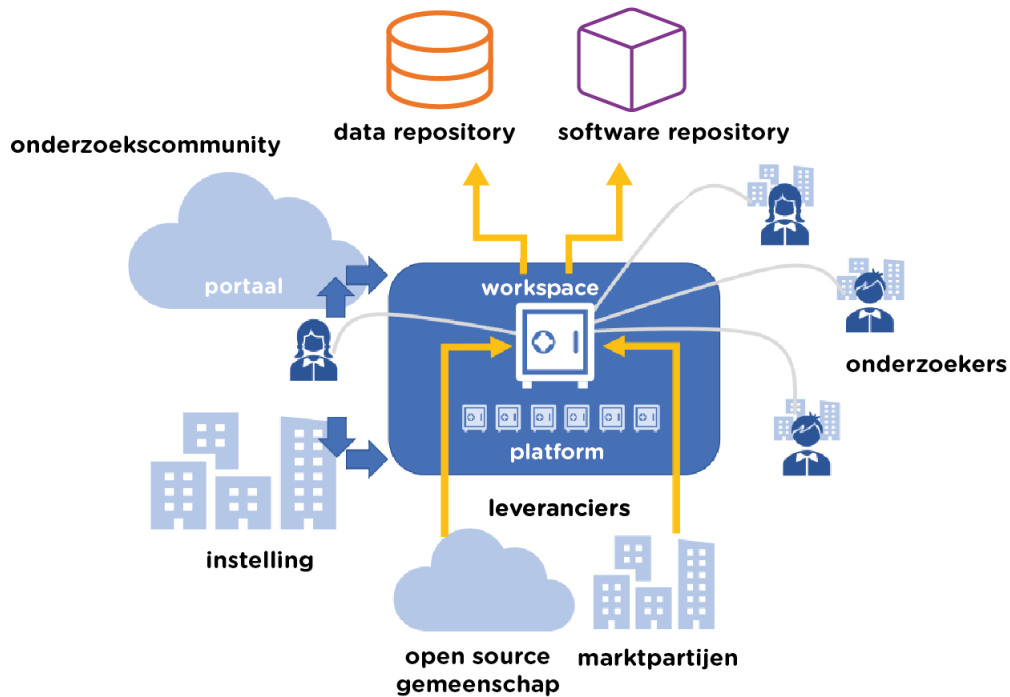
We definiëren de volgende onderdelen van een (hypothetisch) landelijk ecosysteem voor online onderzoeksomgevingen.

- *(VRE-)ecosysteem* = Een systeem waarbinnen in meer of mindere mate wisselwerking mogelijk is tussen (VRE-)portalen, (VRE-)platformen, data-repository’s, software-repository’s en (VRE-) workspaces
 - NB. Een ecosysteem is nadrukkelijk ook een omgeving waar onderzoekinstellingen, marktpartijen et cetera elkaar ontmoeten met een specifiek doel.
- *(VRE-)portaal* = Een digitale/organisatorische thematische (bijvoorbeeld domein, instelling, programma, nationaal, etc.) omgeving, vormgegeven door een actor, zoals een onderzoeksgemeenschap, instelling of DCC, die afnemers toegang geeft tot VRE-workspaces; bijvoorbeeld Health-RI, e/MTIC.
- *(VRE-)platform* = Een technische/functionele/organisatorische omgeving, waarbinnen afnemers een VRE-workspace kunnen gebruiken; bijvoorbeeld anDREa, SURF Research Cloud.
 - NB. Rekenfaciliteiten zullen vaak toegankelijk zijn via de platformen.
- *Data-repository* = Een plek voor opslag, archivering en hergebruik van datasets, vaak met een data-directory geordend naar toepassingsgebied; bijvoorbeeld DANS, 4TU.ResearchData, OpenML.

¹³ HORA: <https://hora.surf.nl/index.php/Hoofdpagina>

¹⁴ ‘Samenwerkingsmodellen voor VRE-waardeketens’, versie 1.0, 12 november 2020

- *Software-repository* = Een plek voor opslag, archivering en hergebruik van scripts, algoritmen, applicaties en andere software, idealiter inclusief versiebeheer, documentatie en metadata; bijvoorbeeld GitLab, OpenML.
- *(VRE-)workspace* = Een digitale plek waar onderzoekers samen kunnen werken aan data-analyse. De workspace biedt toegang tot data en software, en maakt het mogelijk om deze te delen¹⁵

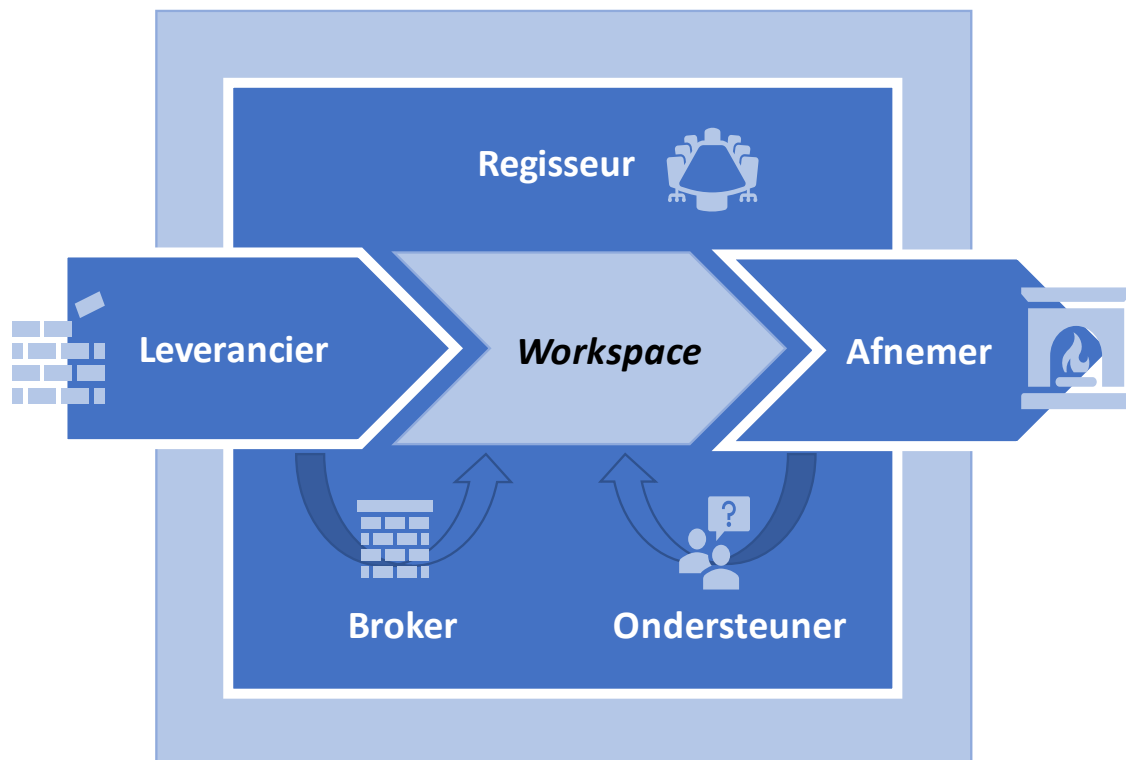


Figuur 3 – Vereenvoudigd overzicht van onderdelen en actoren

¹⁵ Technisch gesproken is de basis van een workspace vaak een zgn. 'container' (bijvoorbeeld Docker of Singularity)

b) Rollen

Een rol is een set aan verantwoordelijkheden, taken en activiteiten. Een rol is een *papieren voorschrift* hoe een actor in die rol zich zou moeten gedragen, dan wel een typering van feitelijk vertoond gedrag ('Organisatie X gedraagt zich in de praktijk als')



Figuur 4 – Schematisch overzicht van de belangrijkste rollen

Voor online onderzoeksomgevingen definiëren we in dit document de volgende rollen (die we omschrijven met een werkwoord om ze goed te onderscheiden van actoren):

- *Afnemer* = Het gebruikmaken van de VRE-workspaces
- *Leverancier* = Het leveren van diensten, componenten en/of infrastructuur voor de VRE-workspaces
- *Regisseur* = Er bestuurlijk voor zorgen dat afnemers en leveranciers goed kunnen functioneren
- *Broker* = Er operationeel voor zorgen dat afnemers diensten, componenten en infrastructuur van leveranciers kunnen gebruiken

- *Ondersteuner* = Het operationeel ondersteunen van de afnemer in het gebruik van de workspace

Een rol kan voorkomen in verschillende onderdelen. Er zouden dus bijvoorbeeld broker-rollen kunnen zijn in een portaal, een platform en een landelijk ecosysteem.

c) Actoren

Een actor is een *fysieke partij*, een groep mensen zoals een organisatie. We onderscheiden *generieke* actoren (groepen, bijvoorbeeld hogescholen) en *specifieke* actoren (individuele partijen, bijvoorbeeld Avans Hogeschool).

Voor online onderzoeksomgevingen definiëren we in dit document de volgende generieke actoren:

- *Onderzoeksgemeenschap's* = Groepen wetenschappers met een gemeenschappelijk onderzoeksgebied; bijvoorbeeld Clariah, DANS, 4TU
- *Allianties* = Groepen van instellingen die een aantal activiteiten gezamenlijk of complementair uitvoeren; bijvoorbeeld TU/e, WUR en UU in The Alliance en UL, TUD en EUR in LDE-Universities
- *Instellingen* = Universiteiten, hogescholen, umc's en onderzoeksinstituten
- *Ict-afdelingen* = Ict-afdelingen, vooral van instellingen
- *DCC's* = Digital Competence Centers, zowel binnen instellingen als zelfstandig, bijvoorbeeld de thematische DCC's
- *SURF* = De landelijke organisatie voor ict in onderzoek en onderwijs
- *Marktpartijen* = Grote en kleine commerciële bedrijven die ict-diensten en -producten leveren, bijvoorbeeld big tech (Microsoft, Google, Amazon), maar ook bijvoorbeeld kleinere lokale bedrijven
- *Open source community's* = Losjes georganiseerde groepen professionals met een gemeenschappelijke interesse, die gezamenlijk ict-diensten en -producten ontwikkelen en beschikbaar stellen

Met medewerking van

- Anke Breeuwsma, Rijksuniversiteit Groningen
- Berry Vetjens, TNO
- Edwin Kallemein, Universiteit van Amsterdam
- Hans van den Berg, Amsterdam UMC
- Henri Hondorp, Universiteit Twente
- Frank Heere, Universiteit Utrecht
- Iman van Lelyveld, De Nederlandse Bank
- Inez Joung, Rijksinstituut voor Volksgezondheid en Milieu
- Irene Nooren, SURF
- Ivar Janmaat, SURF
- Karin van Grieken, SURF
- Louis Stevens, SURF
- Marc Teunis, Hogeschool Utrecht
- Marijke Verheij, Rijksuniversiteit Groningen
- Miaomiao Zhou, Avans Hogeschool
- Michiel Schok, SURF
- Michael Winterkorn, Rijksinstituut voor Volksgezondheid en Milieu
- Monique Uuldriks, Rijksinstituut voor Volksgezondheid en Milieu
- Morris Swertz, UMC Groningen
- Pascal Suppers, Maastricht UMC/Universiteit Maastricht
- Ralph Mettinkhof, Universiteit Twente
- Robert Veen, Erasmus UMC Rotterdam
- Rogier de Jong, SURF
- Ron Augustus, SURF
- Toine Kuiper, TU Eindhoven
- Ymke van den Berg, SURF